

Access DB# 139 601

SEARCH REQUEST FORM
Scientific and Technical Information Center

Requester's Full Name: TANE ZARA Examiner #: 77512 Date: 11-30-04
Art Unit: 1635 Phone Number 302-0765 Serial Number: 07/803,165
Mail Box and Bldg/Room Location: 2728 Results Format Preferred (circle): PAPER DISK E-MAIL
2618

If more than one search is submitted, please prioritize searches in order of need. MEY

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Mutant B types

Inventors (please provide full names): Sobel et al.

Earliest Priority Filing Date: 3-9-01

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search Seq ID No: 34.
AA-774

AA data boxes only. MEY

No size limits

Please search for Interferon
+ regular data
boxes.

MEY

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: _____	NA Sequence (#) _____	STN _____
Searcher Phone #: _____	AA Sequence (#) <u>1</u>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>12/1/04</u>	Bibliographic _____	Dr. Link _____
Date Completed: _____	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems <u>82p</u>
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: _____	Other _____	Other (specify) _____

Thu Dec 2 09:09:45 2004

us-09-803-165-34.rag

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

CM protein - protein search, using sw model

Run on: December 1, 2004, 16:20:27 ; Search time 161 Seconds

(without alignments)
1724.574 Million cell updates/sec

Title: US-09-803-165-34

Perfect score: 4026
Sequence: 1 MIFPDYITDGKFIIRIRK.....KEDLKTQSKQVGLDNLK 774

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 35872929 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

1:	Geneseq1980s:*
2:	Geneseq1950s:*
3:	Geneseq2000s:*
4:	Geneseq2001s:*
5:	Geneseq2002s:*
6:	Geneseq2003as:*
7:	Geneseq2003bs:*
8:	Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4026	100.0	774	4	AAG78941
2	3990	99.1	774	2	AAW29323
3	3840	95.4	774	7	ADG45207
4	3840	95.4	774	7	ADG45209
5	3840	95.4	774	8	ADN75098
6	3840	95.4	774	8	ADN59946
7	3835	95.3	774	7	ADG45210
8	3834	95.2	774	7	ADG45213
9	3834	95.2	774	7	ADG45215
10	3834	95.2	774	7	ADG45240
11	3834	95.2	774	7	ADG45242
12	3834	95.2	774	8	ADG45241
13	3833	95.2	774	8	ADG45241
14	3833	95.2	774	8	ADG45241
15	3832	95.2	774	7	ADG45209
16	3832	95.2	774	7	ADG45210
17	3832	95.2	774	7	ADG45211
18	3832	95.2	774	7	ADG45212
19	3831	95.2	774	7	ADG45212
20	3831	95.2	774	7	ADG45239
21	3831	95.2	774	7	ADG45208
22	3829	95.1	774	8	ADN13792
23	3446.5	85.6	1829	8	AAW29322
24	3446.5	85.6	1829	8	ADN59952
25	3356	83.4	1702	2	AAW38095

26	3253.5	80.8	773	7	ADG45187	Adg45187 Wild type
27	3253.5	80.8	773	7	ADG45187	Adg45187 Wild type
28	3253.5	80.8	773	8	ADN75096	Adn75096 Wild-type
29	3248.5	80.7	773	7	ADG45231	Adg45231 Mutant Tg
30	3248.5	80.7	773	7	ADG45190	Adg45190 Mutant Tg
31	3248.5	80.7	773	8	ADN75105	Adn75105 Tgo DNA p
32	3247.5	80.7	773	7	ADG45195	Adg45195 Mutant Tg
33	3247.5	80.7	773	7	ADG45232	Adg45232 Mutant Tg
34	3247.5	80.7	773	8	ADG45232	Adg45232 Mutant Tg
35	3247.5	80.7	773	8	ADN58201	Adn58201 Mutant Tg
36	3247.5	80.7	773	8	ADN75163	Adn75163 Tgo polym
37	3247.5	80.7	773	8	ADN75164	Adn75164 Tgo polym
38	3246.5	80.6	773	7	ADG45233	Adg45233 Mutant Tg
39	3246.5	80.6	773	8	ADN58200	Adn58200 Mutant Tg
40	3246.5	80.6	773	8	ADN75166	Adn75166 Tgo polym
41	3246.5	80.6	773	8	ADN75165	Adn75165 Tgo polym
42	3246.5	80.6	773	8	ADN75162	Adn75162 Tgo polym
43	3245.5	80.6	773	7	ADG45196	Adg45196 Mutant Tg
44	3245.5	80.6	773	7	ADG45191	Adg45191 Mutant Tg
45	3245.5	80.6	773	7	ADG45189	Adg45189 Mutant Tg

ALIGNMENTS

RESULT 1
AAG78941
ID AAG78941 standard; protein: 774 AA.

AC AAG78941;
DT 13-FEB-2002 (first entry)
XX
DE Recombinant Tag DNA polymerase.
XX
KW Tag polymerase; B-type DNA polymerase.
XX
OS Thermococcus aggregans.
XX
PN Epl132474-A1.
XX
PD 12-SEP-2001.
XX
PF 06-MAR-2001; 2001BP-00104583.
XX
PR 11-MAR-2000; 2000EP-00105155.
XX
RR (HOFF) ROCHE DIAGNOSTICS GMBH.
XX
PA Sobek H, Frey B, Antarakian G, Boehlke K, Pisani FM, Rossi M;
XX
PI WPI; 2001-618367/72.
XX
DR N-PSDB; AAI70053.
XX
PT New thermostable mutant B-type DNA polymerase with a Y-GG/A amino acid
XX motif between the N-terminal 3'-5' exonuclease domain and the C-terminal
XX nucleic acid domain in the wild type polymerase, useful for synthesizing
XX nucleic acids.
XX
PS Disclosure: Fig 9; 40pp; English.
XX
CC The present invention relates to thermostable mutant B-type DNA
XX polymerases, which have a Y-GG/A amino acid motif between the N-terminal
XX 3'-5' exonuclease domain and the C-terminal polymerase domain, where the
XX cytosine of this motif is mutated. The mutant B-type DNA polymerase is
XX useful for synthesizing nucleic acids and for PCR. The present sequence
XX is the protein sequence for a recombinant Thermococcus aggregans (Tag)
XX DNA polymerase, which was used to illustrate the invention
SQ Sequence 774 AA;
Query Match 100.0%; Score 4026; DB 4; Length 774;
Best Local Similarity 100.0%; Pred. No. 4, 2e-305;

Matches 774; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MIPDTDYITKDGKRIIRIFKKENGEPKIEIDPHQPIYALIKDSDAIDEIKAIKGRRG 60
DB 1 MIPDTDYITKDGKRIIRIFKKENGEPKIEIDPHQPIYALIKDSDAIDEIKAIKGRRG 60
QY 61 KIVRVDAVKVKKKFLGRDVEVWKLIFEHPODVPALRGIRHPAVIDIYEDIPAKRY 120
DB 61 KIVRVDAVKVKKKFLGRDVEVWKLIFEHPODVPALRGIRHPAVIDIYEDIPAKRY 120
QY 121 LIDKGLPMEGDEELKMAPIETFFYHGDGFGGKELIMSYADEBARVITWKNIDLPY 180
DB 121 LIDKGLPMEGDEELKMAPIETFFYHGDGFGGKELIMSYADEBARVITWKNIDLPY 180
QY 181 VDVVSNREMIKRFVQIVREKDPDVLITVNGDNFDLPYLKRAEKLGVTLILGRDKEHP 240
DB 181 VDVVSNREMIKRFVQIVREKDPDVLITVNGDNFDLPYLKRAEKLGVTLILGRDKEHP 240
QY 241 PKIHRMGDSFAVEIKGRIFHDLFPVVRRTINLPYTLAEAVYEAULGKTSKLGABEIAAI 300
DB 241 PKIHRMGDSFAVEIKGRIFHDLFPVVRRTINLPYTLAEAVYEAULGKTSKLGABEIAAI 300
QY 301 WETESMKKLAQYSMEDARATYELGKEFFPMEABELAKLIGQSVDSRSSTGNLVEWYLL 360
DB 301 WETESMKKLAQYSMEDARATYELGKEFFPMEABELAKLIGQSVDSRSSTGNLVEWYLL 360
QY 361 RVAVERNELAPNKPDEEYRRRLRTTYLGGYVKEPBERGLMENITYLDFRCIYPSIIVTN 420
DB 361 RVAVERNELAPNKPDEEYRRRLRTTYLGGYVKEPBERGLMENITYLDFRCIYPSIIVTN 420
QY 421 VSPDTLBERGCKNYDVAPIVGYKFCDFPGPIPSILGELITMRQELKKKKATIDPIEKK 480
DB 421 VSPDTLBERGCKNYDVAPIVGYKFCDFPGPIPSILGELITMRQELKKKKATIDPIEKK 480
QY 481 MLDYRQRAVKLANSYGYMGYPKARWYSEKCAESTVANGSHYIEMTIKEIEEFGFKYL 540
DB 481 MLDYRQRAVKLANSYGYMGYPKARWYSEKCAESTVANGSHYIEMTIKEIEEFGFKYL 540
QY 541 YADTDGFYATIPGEKPEITIKKAKEFLKYINSKLPGLLELEGFYLGRGFVAKRYAVI 600
DB 541 YADTDGFYATIPGEKPEITIKKAKEFLKYINSKLPGLLELEGFYLGRGFVAKRYAVI 600
QY 601 DSEGRITTRGLEVRPDSSEIAKETQAKVLEAILKESVEKAVEIVDVEBELAKYQVPL 660
DB 601 DSEGRITTRGLEVRPDSSEIAKETQAKVLEAILKESVEKAVEIVDVEBELAKYQVPL 660
QY 661 EKLVIHQITDLSSEYKAGPHVAIAKRLAKGIVKRGITIIISYIVLRGSGKISDRYILL 720
DB 661 EKLVIHQITDLSSEYKAGPHVAIAKRLAKGIVKRGITIIISYIVLRGSGKISDRYILL 720
QY 721 SEYPPKGRKXDPDYIENQVLPAYLRLIEAFGRKREOLKQSSQVQGLDMLKK 774
DB 721 SEYPPKGRKXDPDYIENQVLPAYLRLIEAFGRKREOLKQSSQVQGLDMLKK 774

```

RESULT 2
 AAM29323
 ID AAM29323 standard; protein: 774 AA.
 XX AAM29323;
 AC 20-APR-1998 (first entry)
 DT
 XX DNA polymerase with 3'-5' exonuclease activity.
 DE
 XX TY Exon: DSM 10597; thermostable; DNA polymerase; 3'-5' exonuclease;
 KM amplification.
 XX Thermococcus sp.
 XX DE19611759-A1.
 XX 02-OCT-1997.

XX 25-MAR-1996; 96DE-01011759.
 XX 25-MAR-1996; 96DE-01011759.
 XX (BOEP) BOEHRINGER MANNHEIM GMBH.
 PI Frey B, Niehaus F, Antmanikien G,
 DR MPI; 1997-481494/45.
 XX N-PSDB; AAT86434.
 PT Thermococcus sp. TY - useful for nucleic
 PT acid amplification.
 PS Claim 1; Page 9-10; 32pp; German.
 CC The present sequence (TY Exon) is a Thermococcus sp. TY (DSM 10597)
 CC thermostable DNA polymerase with 3'-5' exonuclease activity. The enzyme
 CC can specifically amplify nucleic acid fragments of up to 5 kb in high
 CC yields, has an activity half-life of 20 minutes at 90 degrees C, has an
 CC optimum temperature of 70-80 degrees C, has an optimum pH of 7.5,
 CC exhibits optimum activity at a KCl concentration of 80-100 mM, is
 CC magnesium ion-dependent and is inhibited by manganese ions
 SQ Sequence 774 AA;

Query Match 99.1%; Score 3990; DB 2; Length 774;
 Best Local Similarity 99.4%; Pred. No. 2 7e-302;
 Matches 769; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```

QY 1 MIPDTDYITKDGKRIIRIFKKENGEPKIEIDPHQPIYALIKDSDAIDEIKAIKGRRG 60
DB 1 MIPDTDYITKDGKRIIRIFKKENGEPKIEIDPHQPIYALIKDSDAIDEIKAIKGRRG 60
QY 61 KIVRVDAVKVKKKFLGRDVEVWKLIFEHPODVPALRGIRHPAVIDIYEDIPAKRY 120
DB 61 KIVRVDAVKVKKKFLGRDVEVWKLIFEHPODVPALRGIRHPAVIDIYEDIPAKRY 120
QY 121 LIDKGLPMEGDEELKMAPIETFFYHGDGFGKELIMSYADEBARVITWKNIDLPY 180
DB 121 LIDKGLPMEGDEELKMAPIETFFYHGDGFGKELIMSYADEBARVITWKNIDLPY 180
QY 181 VDVVSNREMIKRFVQIVREKDPDVLITVNGDNFDLPYLKRAEKLGVTLILGRDKEHP 240
DB 181 VDVVSNREMIKRFVQIVREKDPDVLITVNGDNFDLPYLKRAEKLGVTLILGRDKEHP 240
QY 241 PKIHRMGDSFAVEIKGRIFHDLFPVVRRTINLPYTLAEAVYEAULGKTSKLGABEIAAI 300
DB 241 PKIHRMGDSFAVEIKGRIFHDLFPVVRRTINLPYTLAEAVYEAULGKTSKLGABEIAAI 300
QY 301 WETESMKKLAQYSMEDARATYELGKEFFPMEABELAKLIGQSVDSRSSTGNLVEWYLL 360
DB 301 WETESMKKLAQYSMEDARATYELGKEFFPMEABELAKLIGQSVDSRSSTGNLVEWYLL 360
QY 361 RVAVERNELAPNKPDEEYRRRLRTTYLGGYVKEPBERGLMENITYLDFRCIYPSIIVTN 420
DB 361 RVAVERNELAPNKPDEEYRRRLRTTYLGGYVKEPBERGLMENITYLDFRCIYPSIIVTN 420
QY 421 VSPDTLBERGCKNYDVAPIVGYKFCDFPGPIPSILGELITMRQELKKKKATIDPIEKK 480
DB 421 VSPDTLBERGCKNYDVAPIVGYKFCDFPGPIPSILGELITMRQELKKKKATIDPIEKK 480
QY 481 MLDYRQRAVKLANSYGYMGYPKARWYSEKCAESTVANGSHYIEMTIKEIEEFGFKYL 540
DB 481 MLDYRQRAVKLANSYGYMGYPKARWYSEKCAESTVANGSHYIEMTIKEIEEFGFKYL 540
QY 541 YADTDGFYATIPGEKPEITIKKAKEFLKYINSKLPGLLELEGFYLGRGFVAKRYAVI 600
DB 541 YADTDGFYATIPGEKPEITIKKAKEFLKYINSKLPGLLELEGFYLGRGFVAKRYAVI 600
QY 601 DSEGRITTRGLEVRPDSSEIAKETQAKVLEAILKESVEKAVEIVDVEBELAKYQVPL 660
DB 601 DSEGRITTRGLEVRPDSSEIAKETQAKVLEAILKESVEKAVEIVDVEBELAKYQVPL 660

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2004, 16:24:27 ; Search time 204 Seconds
(without alignments)
2183.039 Million cell updates/sec

Title: US-09-803-165-34

Perfect score: 4026

Sequence: 1 MIFEDYITKDGKPIIRPK.....KEDLKVQSSKQVGLDPAFLKK 774

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1325181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: 1: uniprot_sprot;*
2: uniprot_trembl;*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	3446.5	85.6	1829 1 DPOL_THERST	033845 thermococcus
2	3356	83.4	1702 1 DPOL_THERLI	P30317 thermococcus
3	3253.5	80.8	773 1 DPOL_THERCO	P56689 thermococcus
4	3212	79.8	775 2 Q9P9K4	Q9P9K4 pyrococcus
5	3211	79.6	775 2 Q9HR06	Q9HR06 pyrococcus
6	3204.5	79.6	775 1 DPOL_THERS9	Q56366 thermococcus
7	3164.5	78.6	771 1 DPOL_PYRAB	P77916 pyrococcus
8	3151	78.3	775 1 DPOL_PYRFU	P61876 pyrococcus
9	3151	78.3	775 1 DPOL_PYRWO	P61876 pyrococcus
10	3141.5	78.0	773 2 Q7S1G7	Q7S1G7 desulfuroc
11	3081	76.5	759 2 Q9HR98	Q9HR98 pyrococcus
12	2934.5	72.9	1312 1 DPOL_PYRSD	Q51134 pyrococcus
13	2905	72.2	733 2 Q7S1G8	Q7S1G8 desulfuroc
14	2886	71.7	1235 1 DPOL_PYRHO	Q59610 pyrococcus
15	2809	69.8	1523 1 DPOL_THERFM	P74918 thermococcus
16	2782	69.1	1671 1 DPOL_PYRKO	P77933 pyrococcus
17	2769.5	68.8	1599 1 DPOL_THERG8	Q9hh84 thermococcus
18	2635.5	63.0	1668 1 DPOL_THERH	Q9hh05 thermococcus
19	1549.5	38.5	784 2 Q6M051	Q6M051 methanococ
20	1549.5	38.5	784 2 CAG2936	CAG2936 methanoc
21	1392.5	34.6	830 2 Q8TWJ5	Q8TWJ5 methanopyr
22	1392	34.6	824 1 DPOL_METVO	P52025 methanocob
23	1301	32.3	781 1 DPOL_ARCFU	Q29753 archaeoglob
24	1292	32.1	1634 1 DPOL_METUA	Q58959 methanococ
25	1209.5	29.9	30.0 DPOL_METHH	Q27276 methanobact
26	1204.5	29.9	781 2 Q9P9N1	Q9P9N1 sulfuriispha
27	1204.5	29.9	781 2 Q96TV1	Q96TV1 sulfobacillu
28	1190	28.6	785 2 Q8ZVW2	Q8ZVW2 pyrobaculum
29	1149.5	28.6	784 1 DPOL_ASRPE	Q93746 aeropyrum p
30	1138	28.3	785 1 Q9P9M2	Q9P9M2 pyrobaculum
31	1125	27.9	785 2 Q9P9M2	Q9P9M2 pyrobaculum

32	914	22.7	933 2 Q9PVG1	Q9PVG1 methanosarc
33	913	22.7	937 2 Q8TSB3	Q8TSB3 methanosarc
34	821	20.4	676 2 Q7AN74	Q7AN74 nanorarchaeu
35	821	20.4	676 2 AAR38923	AAR38923 nanorarcha
36	774	19.2	195 2 Q9HR85	Q9HR85 thermococcus
37	711	17.7	974 2 Q8SQP5	Q8SQP5 encaphalito
38	697.5	17.3	1105 1 DPOL_ORYSA	Q91r66 oryza sativ
39	693.5	17.2	1088 1 DPOL_SOYBN	Q48901 glycine max
40	691.5	17.2	1054 2 Q6BH09	Q6BH09 debaryomyce
41	691	17.2	763 1 DRO3_SULSU	Q05706 sulfobolus
42	676.5	16.8	764 1 DPOL_SULSU	P95979 sulfobolus
43	676.5	16.8	1081 1 DPOL_ARATH	Q91rnt arabidopsis
44	665.5	16.5	1026 2 Q7YVP5	Q7YVP5 typanosoma
45	657.5	16.3	1038 1 DPOL_CANAL	P46588 candida alb

ALIGNMENTS

```

RESULT 1
DPOL_THERST STANDARD; PRT; 1829 AA.
AC Q33845;
AD 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
GN Name=pol;
OS Thermococcus sp. (strain TV).
OC Archaea; Buryarchaeota; Thermococci; Thermococcaceae;
OC Thermococcus.
OX NCBI_TaxID=110163;
RN [1]
RP SOURCE FROM N.A.
RX MEDLINE=96094267; PubMed=9434178;
RA Niehaus F., Frey B., Antikamian G.;
RT "Cloning and characterisation of a thermostable alpha-DNA polymerase
RT from the hyperthermophilic archaeon Thermococcus sp. TV."
RU Gene 204:153-158(1997).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA] (N).
CC -!- PTM: This protein undergoes a protein self splicing that involves
CC a post-translational excision of the three intervening regions
CC (-!- SIMILARITY: Belongs to the DNA polymerase type-B family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y13030; CAA73475.1; -.
CC HSPB; P77933; IGCG.
CC InterPro; IPR006172; DNA_pol_B.
CC InterPro; IPR006133; DNA_pol_B_exo.
CC InterPro; IPR006134; DNA_pol_B_region.
CC InterPro; IPR006142; INTEIN.
CC InterPro; IPR004042; intein_endonuc.
CC InterPro; IPR006141; intein_S.
CC InterPro; IPR004578; pol2.
CC Pfam; PF00136; DNA_pol_B_4.
CC Pfam; PF00136; DNA_pol_B_exo_1.
CC PRINTS; PR00379; INTEIN.
CC TIGRfam; TIGR01443; intein_Cterm; 3.
CC TIGRfam; TIGR01445; intein_Nterm; 3.
CC TIGRfam; TIGR00592; pol2; 2.
CC PROSITE; PS00116; DNA_POLYMERASE_B; FALSE_NEG.
CC PROSITE; PS00818; INTEIN_C_TERM; 3.
CC PROSITE; PS00819; INTEIN_ENDONUCLEASE; 2.
CC PROSITE; PS00817; INTEIN_N_TERM; 3.

```

Autocatalytic cleavage; DNA replication; DNA-binding;
KM DNA-directed DNA polymerase; Endonuclease; Hydrolase;
KM Protein splicing; Transferase.
FT CHAIN 1 409 DNA polymerase, 1st part (Potential).
FT CHAIN 410 769 Intein I.
FT CHAIN 770 855 DNA polymerase, 2nd part (Potential).
FT CHAIN 856 1392 Intein II.
FT CHAIN 1393 1441 DNA polymerase, 3rd part (Potential).
FT CHAIN 1442 1598 Intein III.
FT CHAIN 1599 1829 DNA polymerase, 4th part (Potential).
SQ SEQUENCE 1829 AA; 211875 MW; A113A8BC57E93CB3 CRC64;

Query Match 85.6%; Score 3446.5; DB 1; Length 1829;
Best Local Similarity 42.2%; Pred. No. 9.8e-177;
Matches 771; Conservative 0; Mismatches 3; Indels 1055; Gaps 3;

1 MIFDTYITKDGKPIIRIFKKEGEFKIEBPHFQPIYALAKDSDAIDSIKAKGSRHG 60
1 MILDPTITKDGKPIIRIFKKEGEFKIEBPHFQPIYALAKDSDAIDSIKAKGSRHG 60
61 KIYAVNDAVKKKKFLGRDVAVKLIIEHPDVPAALGKI REHPAVIDIETDIPKRY 120
61 KIYAVNDAVKKKKFLGRDVAVKLIIEHPDVPAALGKI REHPAVIDIETDIPKRY 120
61 KIYAVNDAVKKKKFLGRDVAVKLIIEHPDVPAALGKI REHPAVIDIETDIPKRY 120
121 LIDKGLIPMEGDEELKMAFDIEFTYHGGSEFGKGIIMISYADEBARVITWKNIDLPY 180
121 LIDKGLIPMEGDEELKMAFDIEFTYHGGSEFGKGIIMISYADEBARVITWKNIDLPY 180
121 LIDKGLIPMEGDEELKMAFDIEFTYHGGSEFGKGIIMISYADEBARVITWKNIDLPY 180
181 VDVVSNRERMIKRFQVIVREKDPVLTITNGDNFDEPYLIKRAKLGVTLLGSKDHPE 240
181 VDVVSNRERMIKRFQVIVREKDPVLTITNGDNFDEPYLIKRAKLGVTLLGSKDHPE 240
181 VDVVSNRERMIKRFQVIVREKDPVLTITNGDNFDEPYLIKRAKLGVTLLGSKDHPE 240
241 PKIRHMGSPFAVEIKGRIFHDLFPVVRTINLPYTLAEVAEVLGKTSKGAEIAT 300
241 PKIRHMGSPFAVEIKGRIFHDLFPVVRTINLPYTLAEVAEVLGKTSKGAEIAT 300
241 PKIRHMGSPFAVEIKGRIFHDLFPVVRTINLPYTLAEVAEVLGKTSKGAEIAT 300
301 WETESMKLAQYSMEDARATYELGKEFFMEASHAKLIGSWDVSRSTGNLVEWYL 360
301 WETESMKLAQYSMEDARATYELGKEFFMEASHAKLIGSWDVSRSTGNLVEWYL 360
301 WETESMKLAQYSMEDARATYELGKEFFMEASHAKLIGSWDVSRSTGNLVEWYL 360
361 RYAVERNIAENKPEDEEYRRRLRTTYLGYYKPERGLMENITYLDFRC----- 410
361 RYAVERNIAENKPEDEEYRRRLRTTYLGYYKPERGLMENITYLDFRC----- 410
361 RYAVERNIAENKPEDEEYRRRLRTTYLGYYKPERGLMENITYLDFRC----- 410
411 ----- 410
421 GKGIVNISGVKEGYIIGIDGMQRVKWKVYHSGKLININGLACTPMHKVPVVTENDRQ 480
421 GKGIVNISGVKEGYIIGIDGMQRVKWKVYHSGKLININGLACTPMHKVPVVTENDRQ 480
421 GKGIVNISGVKEGYIIGIDGMQRVKWKVYHSGKLININGLACTPMHKVPVVTENDRQ 480
441 ----- 410
461 TRIRDSLAKSPISGVKRGKITTTKLFKCIAPFKKPKSESEIILGELSGIILAGTILRK 540
461 TRIRDSLAKSPISGVKRGKITTTKLFKCIAPFKKPKSESEIILGELSGIILAGTILRK 540
461 TRIRDSLAKSPISGVKRGKITTTKLFKCIAPFKKPKSESEIILGELSGIILAGTILRK 540
411 ----- 410
541 DIEYDSRGKRIHQRYVEITIGENEKEILERILYLFYLFGRIRSVYKKKGDITNAKI 600
541 DIEYDSRGKRIHQRYVEITIGENEKEILERILYLFYLFGRIRSVYKKKGDITNAKI 600
541 DIEYDSRGKRIHQRYVEITIGENEKEILERILYLFYLFGRIRSVYKKKGDITNAKI 600
411 ----- 410
601 TTAKKAVVLIQIBELLNIESLYAPAVLRGFFERDATTVKRISTIVTQGTNNKKIDIVA 660
601 TTAKKAVVLIQIBELLNIESLYAPAVLRGFFERDATTVKRISTIVTQGTNNKKIDIVA 660
601 TTAKKAVVLIQIBELLNIESLYAPAVLRGFFERDATTVKRISTIVTQGTNNKKIDIVA 660
411 ----- 410
661 KLDSLIGIPYSRYEYKIEENGKELKILITGRDGLIFQTLVGFISSEKNALEKALB 720
661 KLDSLIGIPYSRYEYKIEENGKELKILITGRDGLIFQTLVGFISSEKNALEKALB 720
661 KLDSLIGIPYSRYEYKIEENGKELKILITGRDGLIFQTLVGFISSEKNALEKALB 720
411 ----- 410
721 VRNMRKLKNGSEYKISTFEVSSEYKGVYDLTLEGNPYFANGILTNLSYPSIITYHN 780
721 VRNMRKLKNGSEYKISTFEVSSEYKGVYDLTLEGNPYFANGILTNLSYPSIITYHN 780
721 VRNMRKLKNGSEYKISTFEVSSEYKGVYDLTLEGNPYFANGILTNLSYPSIITYHN 780
421 VSPDTLEEGCKNVDAPIVGYKPKDFPGFIPSTIGLITWQRIKKXKATITDPIKK 480
421 VSPDTLEEGCKNVDAPIVGYKPKDFPGFIPSTIGLITWQRIKKXKATITDPIKK 480
421 VSPDTLEEGCKNVDAPIVGYKPKDFPGFIPSTIGLITWQRIKKXKATITDPIKK 480
761 VSPDTLEEGCKNVDAPIVGYKPKDFPGFIPSTIGLITWQRIKKXKATITDPIKK 840
761 VSPDTLEEGCKNVDAPIVGYKPKDFPGFIPSTIGLITWQRIKKXKATITDPIKK 840
761 VSPDTLEEGCKNVDAPIVGYKPKDFPGFIPSTIGLITWQRIKKXKATITDPIKK 840
491 MLDYQRAVKLHA----- 493

841 MLDYQRAVKLHA----- 493
841 MLDYQRAVKLHA----- 493
841 MLDYQRAVKLHA----- 493
901 VNIFAFSLNKKESKSEIKKVALIRHKYKGEAVEELNSGRKIHITRSHSLFTRNGKI 960
901 VNIFAFSLNKKESKSEIKKVALIRHKYKGEAVEELNSGRKIHITRSHSLFTRNGKI 960
901 VNIFAFSLNKKESKSEIKKVALIRHKYKGEAVEELNSGRKIHITRSHSLFTRNGKI 960
494 ----- 493
494 ----- 493
961 KEIWSGEVKGDLIIVPKKVLNBEKAVINIBELISKLPDEDTADVMTTPVKGKRNPK 1020
961 KEIWSGEVKGDLIIVPKKVLNBEKAVINIBELISKLPDEDTADVMTTPVKGKRNPK 1020
961 KEIWSGEVKGDLIIVPKKVLNBEKAVINIBELISKLPDEDTADVMTTPVKGKRNPK 1020
494 ----- 493
494 ----- 493
1021 GMLRTLKWIFGSESKRIFTFNRYLPHLELAGFVLLPQGEVTDWGLRKYRQLYEKLK 1080
1021 GMLRTLKWIFGSESKRIFTFNRYLPHLELAGFVLLPQGEVTDWGLRKYRQLYEKLK 1080
1021 GMLRTLKWIFGSESKRIFTFNRYLPHLELAGFVLLPQGEVTDWGLRKYRQLYEKLK 1080
494 ----- 493
494 ----- 493
1081 NLRNGNKKELYVRENDIKDSVSCPPKELZEMKIGTKKGPXKCIKRYDEPGKFLGY 1140
1081 NLRNGNKKELYVRENDIKDSVSCPPKELZEMKIGTKKGPXKCIKRYDEPGKFLGY 1140
1081 NLRNGNKKELYVRENDIKDSVSCPPKELZEMKIGTKKGPXKCIKRYDEPGKFLGY 1140
494 ----- 493
494 ----- 493
1141 VSGVAGAKNTGMSYVSKLYNENPVNLKDKNIAEKPFGEVAGKACVDIPKMAVL 1200
1141 VSGVAGAKNTGMSYVSKLYNENPVNLKDKNIAEKPFGEVAGKACVDIPKMAVL 1200
1141 VSGVAGAKNTGMSYVSKLYNENPVNLKDKNIAEKPFGEVAGKACVDIPKMAVL 1200
494 ----- 493
494 ----- 493
1201 LAKSLCGVTAENKRIPIIIFDSSEPVMAFLRAYVGGDIIHPSKRLSTSELIANG 1260
1201 LAKSLCGVTAENKRIPIIIFDSSEPVMAFLRAYVGGDIIHPSKRLSTSELIANG 1260
1201 LAKSLCGVTAENKRIPIIIFDSSEPVMAFLRAYVGGDIIHPSKRLSTSELIANG 1260
494 ----- 493
494 ----- 493
1261 VPLINSIGVSSIKIGPDSGVRYVINEDELPLQTSRKQNTYYPNLIKPEVLEIFGRKE 1320
1261 VPLINSIGVSSIKIGPDSGVRYVINEDELPLQTSRKQNTYYPNLIKPEVLEIFGRKE 1320
1261 VPLINSIGVSSIKIGPDSGVRYVINEDELPLQTSRKQNTYYPNLIKPEVLEIFGRKE 1320
494 ----- 493
494 ----- 493
1321 KNITFEKFEKADSGKLDKRYKLDITLNGDITLDVKNVKEKEGVYDLSVEDNEN 1380
1321 KNITFEKFEKADSGKLDKRYKLDITLNGDITLDVKNVKEKEGVYDLSVEDNEN 1380
1321 KNITFEKFEKADSGKLDKRYKLDITLNGDITLDVKNVKEKEGVYDLSVEDNEN 1380
494 ----- 493
494 ----- 493
1381 FLVGPFLVAHNSYVY 1440
1381 FLVGPFLVAHNSYVY 1440
1381 FLVGPFLVAHNSYVY 1440
543 D----- 543
543 D----- 543
543 D----- 543
1441 DSVTGTDEIIIVKRNRIEFPZIEKLPERVDYRIGEXEYCIILDEVALTLDNKGLIKKV 1500
1441 DSVTGTDEIIIVKRNRIEFPZIEKLPERVDYRIGEXEYCIILDEVALTLDNKGLIKKV 1500
1441 DSVTGTDEIIIVKRNRIEFPZIEKLPERVDYRIGEXEYCIILDEVALTLDNKGLIKKV 1500
544 ----- 543
544 ----- 543
544 ----- 543
1501 PYMRERAKKRYRITITNSWYIDVTEHSLIABDKLEAPMEIBGSLATKODLSG 1560
1501 PYMRERAKKRYRITITNSWYIDVTEHSLIABDKLEAPMEIBGSLATKODLSG 1560
1501 PYMRERAKKRYRITITNSWYIDVTEHSLIABDKLEAPMEIBGSLATKODLSG 1560
544 ----- 543
544 ----- 543
544 ----- 543
1561 VEYIKPEHAIEISYNGVYDIEEGTHRPANGILVHNTDGYATTPGKRPETIKKAKB 1620
1561 VEYIKPEHAIEISYNGVYDIEEGTHRPANGILVHNTDGYATTPGKRPETIKKAKB 1620
1561 VEYIKPEHAIEISYNGVYDIEEGTHRPANGILVHNTDGYATTPGKRPETIKKAKB 1620
544 ----- 543
544 ----- 543
544 ----- 543
566 FLKYINSKLPGLLELYEGFYLRGFFVAKKRAYVDESGRIITRGLVVRBMSIAXET 625
566 FLKYINSKLPGLLELYEGFYLRGFFVAKKRAYVDESGRIITRGLVVRBMSIAXET 625
566 FLKYINSKLPGLLELYEGFYLRGFFVAKKRAYVDESGRIITRGLVVRBMSIAXET 625
1621 FLKYINSKLPGLLELYEGFYLRGFFVAKKRAYVDESGRIITRGLVVRBMSIAXET 1680
1621 FLKYINSKLPGLLELYEGFYLRGFFVAKKRAYVDESGRIITRGLVVRBMSIAXET 1680
1621 FLKYINSKLPGLLELYEGFYLRGFFVAKKRAYVDESGRIITRGLVVRBMSIAXET 1680
494 ----- 493
494 ----- 493
626 QAVTLEALIKEDSVEKAVIIVDVVEIANKQVPLEKVIHROITQDLSFYAIGSHVAI 685
626 QAVTLEALIKEDSVEKAVIIVDVVEIANKQVPLEKVIHROITQDLSFYAIGSHVAI 685
626 QAVTLEALIKEDSVEKAVIIVDVVEIANKQVPLEKVIHROITQDLSFYAIGSHVAI 685
1681 QAVTLEALIKEDSVEKAVIIVDVVEIANKQVPLEKVIHROITQDLSFYAIGSHVAI 1740
1681 QAVTLEALIKEDSVEKAVIIVDVVEIANKQVPLEKVIHROITQDLSFYAIGSHVAI 1740
1681 QAVTLEALIKEDSVEKAVIIVDVVEIANKQVPLEKVIHROITQDLSFYAIGSHVAI 1740
494 ----- 493
494 ----- 493
686 AKELAAKGIKVRGTLIISTYVLRSGKISDRVILSEYDPKKGKHYDPDYIENQVLPAYL 745
686 AKELAAKGIKVRGTLIISTYVLRSGKISDRVILSEYDPKKGKHYDPDYIENQVLPAYL 745
686 AKELAAKGIKVRGTLIISTYVLRSGKISDRVILSEYDPKKGKHYDPDYIENQVLPAYL 745
1741 AKELAAKGIKVRGTLIISTYVLRSGKISDRVILSEYDPKKGKHYDPDYIENQVLPAYL 1800
1741 AKELAAKGIKVRGTLIISTYVLRSGKISDRVILSEYDPKKGKHYDPDYIENQVLPAYL 1800
1741 AKELAAKGIKVRGTLIISTYVLRSGKISDRVILSEYDPKKGKHYDPDYIENQVLPAYL 1800
494 ----- 493
494 ----- 493
746 RILEAFGRKEDLKYSKQVGLDAMLKK 774
746 RILEAFGRKEDLKYSKQVGLDAMLKK 774
746 RILEAFGRKEDLKYSKQVGLDAMLKK 774
1801 RILEAFGRKEDLKYSKQVGLDAMLKK 1829
1801 RILEAFGRKEDLKYSKQVGLDAMLKK 1829
1801 RILEAFGRKEDLKYSKQVGLDAMLKK 1829

RESULT 2
DPOL_THELI
ID DPOL_THELI STANDARD; PRT; 1702 AA.
AC P30317;
DT 01-APR-1993 (Rel. 25, Created)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2004, 16:23:42 ; Search time 21 Seconds

(without alignments)
3546.275 Million cell updates/sec

Title: US-09-803-165-34

Sequence: 1 MIFPTDYITKDGKPIIRIFK.....KEDIKYQSSKQVGDAMLKK 774

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: PIR 79:*

1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3356	83.4	1702	2	542459 DNA-directed DNA p
2	3204.5	79.6	775	2	567920 DNA-directed DNA p
3	3164.5	78.6	771	2	575023 DNA polymerase I p
4	3151	78.3	775	2	535543 DNA-directed DNA p
5	2934.5	72.9	1312	2	568593 DNA-directed DNA p
6	2886	71.7	1235	2	671210 probable DNA-direc
7	2764.5	68.7	1670	2	571551 DNA-directed DNA p
8	1301	32.3	781	2	569312 DNA polymerase BI
9	1291	32.1	1634	2	564410 DNA-directed DNA p
10	1209.5	30.0	586	2	569028 DNA-dependent DNA
11	1204.5	29.9	781	2	567382 DNA-directed DNA p
12	1190	29.6	803	2	562777 probable DNA-direc
13	1138	28.3	784	2	572515 DNA-directed DNA p
14	693.5	17.2	1088	2	575407 probable DNA-direc
15	676.5	16.8	764	2	575407 probable DNA-direc
16	659.5	16.4	1038	2	575407 probable DNA-direc
17	657.5	16.3	1038	2	575407 probable DNA-direc
18	657.5	16.3	1038	2	575407 probable DNA-direc
19	649.5	16.3	1086	2	575407 probable DNA-direc
20	626	15.5	901	2	575407 probable DNA-direc
21	621.5	15.4	1097	1	575407 probable DNA-direc
22	619.5	15.4	1084	1	575407 probable DNA-direc
23	619.5	15.4	1084	1	575407 probable DNA-direc
24	616.5	15.3	872	2	575407 probable DNA-direc
25	616.5	15.3	872	2	575407 probable DNA-direc
26	616.5	15.3	1105	1	575407 probable DNA-direc
27	614.5	15.2	1094	1	575407 probable DNA-direc
28	611	15.2	1107	1	575407 probable DNA-direc
29	609	15.1	1513	2	575407 probable DNA-direc

30	605.5	15.0	1106	1	A39299 DNA-directed DNA p
31	587	14.6	1492	2	T18560 DNA-directed DNA p
32	584.5	14.5	879	2	A56277 DNA-directed DNA p
33	583.5	14.5	1462	1	DJHMAC DNA-directed DNA p
34	581	14.4	1081	2	T20698 DNA-directed DNA p
35	578.5	14.4	875	2	UC5186 DNA-directed DNA p
36	570.5	14.2	959	2	F72763 probable DNA-direc
37	570	14.2	1465	2	S45628 DNA-directed DNA p
38	559.5	13.9	223	2	E69125 DNA-directed DNA p
39	548.5	13.6	1339	1	S20052 DNA-directed DNA p
40	527.5	13.1	1468	2	S58250 DNA-directed DNA p
41	522.5	13.0	1505	2	S28079 DNA-directed DNA p
42	516.5	12.8	844	2	T31321 DNA-directed DNA p
43	516	12.5	1405	1	DJZDA DNA-directed DNA p
44	504.5	12.5	2154	2	A84669 DNA-directed DNA p
45	500.5	12.4	913	2	T17675 DNA-directed DNA p

ALIGNMENTS

RESULT 1

S42459
DNA-directed DNA polymerase (EC 2.7.7.7) Vent, intein containing precursor - Thermococcus
NContains: DNA endonuclease (EC 3.1.1.1) PI-I; DNA endonuclease (EC 3.1.1.1) PI-II; DN
C:Species: Thermococcus litoralis
C:Date: 31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change 09-Jul-2004
C:Accession: S42459; S42451; S42450; S42458
R:Perler, F.B.; Comb, D.G.; Jack, W.E.; Moran, L.S.; Qiang, B.; Kucera, R.B.; Benner, J
submitted to the EMBL Data Library, September 1992
A:Reference number: S42458
A:Accession: S42459
A:Molecule type: DNA
A:Residues: 1-1702 (PERI>
A:Cross-references: UNIPROT:P30317; EMBL:W4198; NID:G154685; PIDD:AAA72100.1; PIDD:154
R:Perler, F.B.; Comb, D.G.; Jack, W.E.; Moran, L.S.; Qiang, B.; Kucera, R.B.; Benner, J
Proc. Natl. Acad. Sci. U.S.A. 89, 5577-5581, 1992
A:Title: Intervening sequences in an Archaea DNA polymerase gene.
A:Reference number: S42450; WUID:92302285; PMID:1608969
A:Accession: S42451
A:Molecule type: DNA
A:Residues: 181-222;387-425;452-476;483-524;1021-1062;1076-1099;1466-1489;1533-1547 <PE
A:Cross-references: EMBL:W4198
R:Hodges, R.A.; Perler, F.B.; Noren, C.J.; Jack, W.E.
Nucleic Acids Res. 20, 6153-6157, 1992
A:Title: Protein splicing removes intervening sequences in an archaea DNA polymerase.
A:Reference number: S40788; WUID:93117083; PMID:1475179
A:Contents: annotation
C:Function: <VENT>
A:Description: nucleotidyltransferase
A:Note: DNA-directed DNA polymerase Vent
C:Function: <ENDI>
A:Description: endonuclease; hydrolase
A:Note: DNA endonuclease PI-TII
C:Function: <END2>
A:Description: endonuclease; hydrolase
A:Note: DNA endonuclease PI-TII
C:Superfamily: Thermococcus litoralis DNA-directed DNA polymerase Vent
C:Keywords: DNA binding; endonuclease; hydrolase; nucleotidyltransferase; protein splic
F:1-494;1033-1081;1472-1702/Product: DNA-directed DNA polymerase Vent #status predicted <X1>
F:1-494/Domain: DNA-directed DNA polymerase Vent extein 1 #status predicted <X1>
F:495-1032/Product: DNA-directed DNA polymerase PI-II (pol Vent intein 1) #status predicted <MAT2>
F:1033-1081/Domain: DNA-directed DNA polymerase Vent intein 2 #status predicted <MAT2>
F:1082-1471/Product: DNA endonuclease PI-I (pol Vent intein 2) #status predicted <X1>
F:1472-1702/Domain: DNA-directed DNA polymerase Vent extein 3 #status predicted <MAT3>
F:1494-1033/Cross-link: peptide (Asn-Ser) #status predicted
F:1081-1472/Cross-link: peptide (Asp-Thr) #status predicted

Query Match 83.4%; Score 3356; DB 2; Length 1702;
Best Local Similarity 42.7%; Pred. No. 3; 8e-166;
Matches 726; Conservative 33; Mismatches 15; Indels 928; Gaps 2;
1 MIFPTDYITKDGKPIIRIFKENGKFEKLEDPHFQYITVALKQDSAIKIGRRHG 60

```

Db 1 MIDDTDTITDKGPRITIRIKKENGFEKILDPHPQPIYALIKDSDAIBELIKAIKBERRG 60
QY KIVRVDAVAVKVKKFLGRDVEVWKLIFEHQODVPALRGKIREHPAVIDIYEYDIPAKKY 120
Db 61 KIVRVDAVAVKVKKFLGRDVEVWKLIFEHQODVPALRGKIREHPAVIDIYEYDIPAKKY 120
QY LIDGGLIPMGDEELKLMADIFETFEHGEDEFGKGIIMISYADEBEARVITMKNIDLPY 180
Db 121 LIDGGLIPMGDEELKLMADIFETFEHGEDEFGKGIIMISYADEBEARVITMKNIDLPY 180
QY 181 VDVSNEREMIKRFVQIVREKDPVLLITNGDNFDLPYLTKAEKGLVTLGDKHEPE 240
Db 181 VDVSNEREMIKRFVQIVREKDPVLLITNGDNFDLPYLTKAEKGLVTLGDKHEPE 240
QY 241 PKIRMGDSFAVELIKGRIHFDLPFVARTINLPYTLAEVYAVIGKTKKGLAEBELAI 300
Db 241 PKIRMGDSFAVELIKGRIHFDLPFVARTINLPYTLAEVYAVIGKTKKGLAEBELAI 300
QY 301 WEITESMKKLAQYSMEDARATYSLGKEPFPEKAEALAKLIGOSVMDVSRSSGTGLVEMYLL 360
Db 301 WEITESMKKLAQYSMEDARATYSLGKEPFPEKAEALAKLIGOSVMDVSRSSGTGLVEMYLL 360
QY 361 RVAERHBIAPNKDEDEEYRRRLRTTYLGIVYKPEPERGLMNTIYLDPRCISYIIVYHN 420
Db 361 RVAERHBIAPNKDEDEEYRRRLRTTYLGIVYKPEPERGLMNTIYLDPRCISYIIVYHN 420
QY 421 VSPDLTEREGCKNDVAPIVGKFCDFPGFIPSIILGELLITMGEIKKKKNAIIPDIIEKK 480
Db 421 VSPDLTEREGCKNDVAPIVGKFCDFPGFIPSIILGELLITMGEIKKKKNAIIPDIIEKK 480
QY 481 MLADYRQAVKLAHA----- 493
Db 481 MLADYRQAVKLAHA----- 493
QY 494 ----- 493
Db 494 ----- 493
QY 541 VNNLFAPSEFKKIKESSEYKVKYKALIRHKYKSKAYEIQSSGRKINITAGSLFTVRNGEI 600
Db 541 VNNLFAPSEFKKIKESSEYKVKYKALIRHKYKSKAYEIQSSGRKINITAGSLFTVRNGEI 600
QY 494 ----- 493
Db 494 ----- 493
QY 601 KEVSGDGIKEGDLIAPKIKILNEKVSINIPBLISDSEETADIVMTISAKGRKAPFK 660
Db 601 KEVSGDGIKEGDLIAPKIKILNEKVSINIPBLISDSEETADIVMTISAKGRKAPFK 660
QY 494 ----- 493
Db 494 ----- 493
QY 661 GMLRTLRMPGSENRIRITFNRYLPHLEKGLIKLIPRGYVTDWELRKYKQLYEKLAK 720
Db 661 GMLRTLRMPGSENRIRITFNRYLPHLEKGLIKLIPRGYVTDWELRKYKQLYEKLAK 720
QY 494 ----- 493
Db 494 ----- 493
QY 721 SVKYNNGKREYLVFNELIKDISYFPOKELEBWKIGTLNGFRINCILKVDDEFGKLLGY 780
Db 721 SVKYNNGKREYLVFNELIKDISYFPOKELEBWKIGTLNGFRINCILKVDDEFGKLLGY 780
QY 494 ----- 493
Db 494 ----- 493
QY 781 VSEGYAGQKNKKTGISYSVTLNEDPVVLESMDKVAEKFQKRVDRNCVISKQAVYL 840
Db 781 VSEGYAGQKNKKTGISYSVTLNEDPVVLESMDKVAEKFQKRVDRNCVISKQAVYL 840
QY 494 ----- 493
Db 494 ----- 493
QY 901 VFLNLSIGISSVKIGFDSGVRYVINEDLPQPTSRKENTYYSNLIPKELIADVFGKRFQ 960
Db 901 VFLNLSIGISSVKIGFDSGVRYVINEDLPQPTSRKENTYYSNLIPKELIADVFGKRFQ 960
QY 494 ----- 493
Db 494 ----- 493
QY 961 KNNTFKFKKELVDSGLNREKAKLLEFFINGDILDRYKSVKEXDVEGYVYDLSVEDNEN 1020
Db 961 KNNTFKFKKELVDSGLNREKAKLLEFFINGDILDRYKSVKEXDVEGYVYDLSVEDNEN 1020
QY 494 ----- 542
Db 494 ----- 542
QY 1021 FLVNGGLIYAHNSYTGWGPYKARVYSKECAESVTAMGRHITMTIRIRIEKPFKLYLA 1080
Db 1021 FLVNGGLIYAHNSYTGWGPYKARVYSKECAESVTAMGRHITMTIRIRIEKPFKLYLA 1080
QY 543 D----- 543

```

```

Db 1081 DSVSESEIIIRONGKIRFVKIKDLFSKVDSIGEKEYCILEGEVALTLDDGKLWMPV 1140
QY 544 ----- 543
Db 1141 PYVMEHRANKRMFRIMLNSWYIDVTEDEHSLIGYINTSKTKAKKIGBLKEVFPBLGK 1200
QY 544 ----- 543
Db 1201 AVKSLICPNAPLKDENTKTSIAVKEFWELVGLVGDQNGGDSRWABYYLGLSTGDAEE 1260
QY 544 ----- 543
Db 1261 IKOKLEPLKTYGVISNTYPPAKKEDFNILAKSLVFKPKRHPKDKGRKKIPEFMYELPV 1320
QY 544 ----- 543
Db 1321 TYIEAFRLGLFSADGTVIRKGVPEIRLTNIDADFLREVRLMTVGISNSIFAETPNR 1380
QY 544 ----- 543
Db 1381 YNGVSTGYTSKHLRIKNNKRFARIGFLIERQKALLEHKSARVAKNTIDGFDLVHYK 1440
QY 544 ----- 1441
Db 1441 KYEBIPYEGVYDIEVESTHRFFANNILVHNTDGFATIPGKPELIKKAKEFLMYINS 1500
QY 573 KLPGLLEFEGEPYLRGFPVAKKRYAVIDEBGRIITRGLEVYRPMSEIAKEQAKVLEA 632
Db 1501 KLPGLLEFEGEPYLRGFPVAKKRYAVIDEBGRIITRGLEVYRPMSEIAKEQAKVLEA 632
QY 633 ILKEDSVEKAVIIVKDVAEELAKYQVPLEKLVHBOITKDLEBYAIGPHVALAKRLAAK 692
Db 1561 ILKESVEKAVAVDVAVEKIKAYRPLEKLVHBOITKDLEBYAIGPHVALAKRLAAK 692
QY 693 GIKRFGTIIISTYVLRGSGKISDRVITLSEYQPKKXNDPVYIENOVYPAVRLITAEAG 752
Db 1621 GIKRFGTIIISTYVLRGSGKISDRVITLSEYQPKKXNDPVYIENOVYPAVRLITAEAG 752
QY 753 YRKEDLKYSKQVGLDAMIKK 774
Db 1681 YRKEDLKYSKQVGLDAMIKK 774

```

RESULT 2
 S67920
 DNA-directed DNA polymerase (EC 2.7.7.7) - Thermococcus sp.
 C:Species: Thermococcus sp.
 C:Date: 17-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 09-Jul-2004
 C:Accession: S67920
 R:Southworth, M.S.; Kong, H.; Kucera, R.B.; Ware, J.; Jannasch, H.W.; Pelzer, F.B.
 Submitted to the EMBL Data Library, January 1996
 A:Description: Cloning, expression and modulation of the 3'-5' exonuclease activity of
 A:Reference number: S67920
 A:Accession: S67920
 A:Molecule type: DNA
 A:Residues: 1-775 <SOU>
 A:Cross-references: UNIPROT:Q5366; GB:U47108; NID:g1197451; PIDD:AAA8769.1; PID:g1197
 A:Experimental source: strain 90N-7
 C:Superfamily: DNA polymerase
 C:Keywords: DNA binding; nucleotidyltransferase

Query Match 79.6%; Score 3204.5; DB 2; Length 775;
Best local similarity 76.6%; Pred. No. 9,4e-159;
Matches 592; Conservative 89; Mismatches 89; Indels 3; Gaps 2;

```

QY 1 MIDDTDTITDKGPRITIRIKKENGFEKILDPHPQPIYALIKDSDAIBELIKAIKBERRG 60
Db 1 MIDDTDTITDKGPRITIRIKKENGFEKILDPHPQPIYALIKDSDAIBELIKAIKBERRG 60
QY 61 KIVRVDAVAVKVKKFLGRDVEVWKLIFEHQODVPALRGKIREHPAVIDIYEYDIPAKKY 120
Db 61 KIVRVDAVAVKVKKFLGRDVEVWKLIFEHQODVPALRGKIREHPAVIDIYEYDIPAKKY 120
QY 121 LIDGGLIPMGDEELKLMADIFETFEHGEDEFGKGIIMISYADEBEARVITMKNIDLPY 180
Db 121 LIDGGLIPMGDEELKLMADIFETFEHGEDEFGKGIIMISYADEBEARVITMKNIDLPY 180
QY 181 VDVSNEREMIKRFVQIVREKDPVLLITNGDNFDLPYLTKAEKGLVTLGDKHEPE 240
Db 181 VDVSNEREMIKRFVQIVREKDPVLLITNGDNFDLPYLTKAEKGLVTLGDKHEPE 240
QY 241 PKIRMGDSFAVELIKGRIHFDLPFVARTINLPYTLAEVYAVIGKTKKGLAEBELAI 300
Db 241 PKIRMGDSFAVELIKGRIHFDLPFVARTINLPYTLAEVYAVIGKTKKGLAEBELAI 300
QY 301 WEITESMKKLAQYSMEDARATYSLGKEPFPEKAEALAKLIGOSVMDVSRSSGTGLVEMYLL 360
Db 301 WEITESMKKLAQYSMEDARATYSLGKEPFPEKAEALAKLIGOSVMDVSRSSGTGLVEMYLL 360
QY 361 RVAERHBIAPNKDEDEEYRRRLRTTYLGIVYKPEPERGLMNTIYLDPRCISYIIVYHN 420
Db 361 RVAERHBIAPNKDEDEEYRRRLRTTYLGIVYKPEPERGLMNTIYLDPRCISYIIVYHN 420
QY 421 VSPDLTEREGCKNDVAPIVGKFCDFPGFIPSIILGELLITMGEIKKKKNAIIPDIIEKK 480
Db 421 VSPDLTEREGCKNDVAPIVGKFCDFPGFIPSIILGELLITMGEIKKKKNAIIPDIIEKK 480
QY 481 MLADYRQAVKLAHA----- 493
Db 481 MLADYRQAVKLAHA----- 493
QY 494 ----- 493
Db 494 ----- 493
QY 541 VNNLFAPSEFKKIKESSEYKVKYKALIRHKYKSKAYEIQSSGRKINITAGSLFTVRNGEI 600
Db 541 VNNLFAPSEFKKIKESSEYKVKYKALIRHKYKSKAYEIQSSGRKINITAGSLFTVRNGEI 600
QY 494 ----- 493
Db 494 ----- 493
QY 601 KEVSGDGIKEGDLIAPKIKILNEKVSINIPBLISDSEETADIVMTISAKGRKAPFK 660
Db 601 KEVSGDGIKEGDLIAPKIKILNEKVSINIPBLISDSEETADIVMTISAKGRKAPFK 660
QY 494 ----- 493
Db 494 ----- 493
QY 661 GMLRTLRMPGSENRIRITFNRYLPHLEKGLIKLIPRGYVTDWELRKYKQLYEKLAK 720
Db 661 GMLRTLRMPGSENRIRITFNRYLPHLEKGLIKLIPRGYVTDWELRKYKQLYEKLAK 720
QY 494 ----- 493
Db 494 ----- 493
QY 721 SVKYNNGKREYLVFNELIKDISYFPOKELEBWKIGTLNGFRINCILKVDDEFGKLLGY 780
Db 721 SVKYNNGKREYLVFNELIKDISYFPOKELEBWKIGTLNGFRINCILKVDDEFGKLLGY 780
QY 494 ----- 493
Db 494 ----- 493
QY 781 VSEGYAGQKNKKTGISYSVTLNEDPVVLESMDKVAEKFQKRVDRNCVISKQAVYL 840
Db 781 VSEGYAGQKNKKTGISYSVTLNEDPVVLESMDKVAEKFQKRVDRNCVISKQAVYL 840
QY 494 ----- 493
Db 494 ----- 493
QY 901 VFLNLSIGISSVKIGFDSGVRYVINEDLPQPTSRKENTYYSNLIPKELIADVFGKRFQ 960
Db 901 VFLNLSIGISSVKIGFDSGVRYVINEDLPQPTSRKENTYYSNLIPKELIADVFGKRFQ 960
QY 494 ----- 493
Db 494 ----- 493
QY 961 KNNTFKFKKELVDSGLNREKAKLLEFFINGDILDRYKSVKEXDVEGYVYDLSVEDNEN 1020
Db 961 KNNTFKFKKELVDSGLNREKAKLLEFFINGDILDRYKSVKEXDVEGYVYDLSVEDNEN 1020
QY 494 ----- 542
Db 494 ----- 542
QY 1021 FLVNGGLIYAHNSYTGWGPYKARVYSKECAESVTAMGRHITMTIRIRIEKPFKLYLA 1080
Db 1021 FLVNGGLIYAHNSYTGWGPYKARVYSKECAESVTAMGRHITMTIRIRIEKPFKLYLA 1080
QY 543 D----- 543

```